REVOLUTIONIZING BIOMEDICAL RESEARCH: INTEGRATING CUTTING-EDGE AI/ML TO UNLEASH INNOVATION IN DRUG DISCOVERY AND THERAPEUTICS DEVELOPMENT

Conotoxin Prediction And Classification: New Features To Increase Prediction Accuracy

Hau Nguyen Los Alamos National Laboratory **Hau Nguyen** Los Alamos National Laboratory **Duc Truong** Los Alamos National Laboratory **Lyman Monroe** Los Alamos National Laboratory **Robert Williams** Los Alamos National Laboratory

Conotoxins are small molecule toxins isolated from the venom of marine cone snails, which use these neurotoxic peptides to paralyze their prey. Conotoxins pose a significant threat to humans because they target multiple receptors and ion channels, causing a wide range of pathophysiological effects that can lead to incapacitation and/or death. At the same time, they also have extraordinary potential for medical therapeutics that include cancer, microbial infections, epilepsy, autoimmune diseases, neurological conditions, and cardiovascular disorders. Despite the critical roles of conotoxins in biodefense and their strong potentials in novel therapeutic treatment development, the process of identifying and characterizing the toxicities of conotoxins and similar small molecule toxins is difficult, costly, and time-consuming, which often requires a series of diverse and complex biological, toxicological, and analytical techniques for effective characterization.

Recent machine learning (ML) models, trained solely on primary amino acid sequences to predict biological toxins (e.g. conotoxins and animal venoms) suffer from many limitations due to peptide conformational flexibility and the high frequency of cysteines resulting in an enumerable set of disulfide-bridged foldamers. The same primary amino acid sequence can produce multiple molecules with different conformations that affect their function and toxicity levels. A given peptide may be toxic when its cysteine residues form a particular disulfide-bond pattern, while alternative bonding patterns (isoforms) or its reduced form (free cysteines with no disulfide bridges) may have little or no toxicological effects. Similarly, the same disulfide-bond pattern may be possible for other peptide sequences and result in different conformations that all exhibit varying toxicities to the same receptor or to different receptors. We present here new features, when combined with primary sequence features to train ML algorithms to predict conotoxins, that significantly increase prediction accuracy, both in predicting conotoxins and conotoxin classes. This platform not only accelerates the identification of novel chemical and biochemical threat agents but also benefits biological prophylaxes/therapeutics and the development of medical countermeasures.

This work is released for publication in accordance with LANL LA-UR-24-24381 by Triad National Security, LLC operator of the Los Alamos National Laboratory under contract no. 89233218CNA000001 with the U.S. Department of Energy.